Source and Accuracy of Estimates

SOURCES OF DATA

Most estimates in this report come from data obtained in March of years 1968 through 2001 in the Current Population Survey (CPS). The Census Bureau conducts the survey every month, although this report uses only March data for its estimates. The March survey uses two sets of questions, the basic CPS and the supplement.

The Census Bureau used data from various sources in developing alternative measures of income for 2000. Specifically, we combined data from the American Housing Survey (AHS), the Income Survey Development Program (ISDP), and the Internal Revenue Service (IRS) with CPS data to create simulations of taxes paid, number of tax filing units, adjusted gross income, and other tax characteristics for the March 2001 CPS.

In addition, this report uses the *State Tax Handbook* from the Commerce Clearing House as an information source for tax data. For noncash valuation estimates, this report uses data from the U.S. Department of Agriculture (USDA), the Health Care Financing Administration (HCFA), and the Department of Housing and Urban Development (HUD).

A description of the sources of data we used to derive these estimates follows. Except for the CPS, these descriptions are brief. See Current Population Reports, Series P60-186RD, *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992,* and publications on the appropriate surveys for more details.

American Housing Survey. The Census Bureau collects housing data for the Department of Housing and Urban Development. The population covered by the sample for the AHS (called the Annual Housing Survey before 1984) includes all housing units in the United States. For a more detailed description of the sample design, see the report Current Housing Reports, Series H150-89, *The American Housing Survey for the United States in 1989,* U.S. Department of Commerce.

The AHS is no longer conducted in even-numbered years, so we based the property tax estimates in this report on the 1995 AHS. Also, for the noncash estimates, we used the 1985 AHS data in a model to estimate the value of public housing. For more details on the AHS model used to estimate public and subsidized housing values, please see Appendix B of Current Population Reports, Series P60-186RD, *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992*.

Income Survey Development Program. The ISDP was the research and development phase for the Survey of Income and Program Participation (SIPP). The Census Bureau used the ISDP to examine and resolve design, operational, and technical issues for SIPP. The household sample for the 1979 ISDP was a nationwide, multiple-frame sample. The majority of sample households in the ISDP came from addresses contacted in the 1976 Survey of Income and Education. Statisticians selected the remainder of sample households from a reserve file of sample cases maintained by the Census Bureau. For a more detailed description of this sample design, see the report *Wage and Salary Data From the Income Survey Development Program: 1979 (Preliminary Data From Interview Period One),* Current Population Reports, Special Studies, Series P-23, No. 118.

Internal Revenue Service data. Much of the IRS data in this report come from the Statistics of Income (SOI) series, in particular the SOI Bulletin *Individual Income Tax Returns, Preliminary Data: 1999,* Spring 2001. This report, based on a sample drawn from all tax returns filed in 2000, presents information on taxpayers' incomes, exemptions, deductions, credits, and taxes.

Data from other sources. The *State Tax Handbook,* October 1, 1991, from the Commerce Clearing House, includes information on state tax systems. We updated these data to reflect changes in state income tax rates.

Much of the data on cash and noncash benefits are from administrative records. Values of school lunches and food stamps are from USDA unpublished data. Medicaid and medicare data come from HCFA unpublished records. Also, USDA and HUD data are used to compute medicaid and medicare values. For more details, see Appendix B of Current Population Reports, Series P60-186RD, *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992*.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Field representatives ask questions concerning labor force participation about each member 15 years old and over in every sample household.

The CPS sample includes coverage in all 50 states and the District of Columbia. The Census Bureau continually updates the sample to account for new residential construction. The Census Bureau divides the United States into 2,007 geographic areas. In most states, a

geographic area consists of a county or several contiguous counties. In some areas of New England and Hawaii, the Census Bureau uses minor civil divisions instead of counties. We select a total of 754 geographic areas for sample. About 50,000 occupied households are eligible for interview every month. Field representatives are unable to obtain interviews at about 3,200 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Census Bureau has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The Census Bureau completely implemented the most recent changes due to the 1990 census-based redesign in July 1995.

Table E–1 summarizes changes in the CPS designs for the years for which data appear in this report.

CPS March supplement. In addition to the basic CPS questions, field representatives asked supplementary questions in March about money income received the previous calendar year.

To obtain more reliable data for the Hispanic-origin population, the Census Bureau increased the March CPS sample by about 2,500 eligible housing units, interviewed the previous November, that contained at least one sample person of Hispanic origin.[1] In addition, the sample included people in the armed forces living off post or with their families on post.

CPS estimation procedure. This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by state, age, sex, race, and Hispanic/non-Hispanic categories.

The independent estimates are based on:

- The 1990 Decennial Census of Population and Housing.
- An adjustment for undercoverage in the 1990 census.
- Statistics on births, deaths, immigration, and emigration.
- Statistics on the size of the armed forces.

The independent population estimates used for 1994 (1993 for income estimates) and later are based on updates to controls established by the 1990 decennial census. Data previous to 1994 are based on independent population estimates from the latest available decennial census data. The estimation procedure for the March supplement included a further adjustment so husband and wife of a household received the same weight.

Table E–1. **Description of the March Current Population Survey**

Time period	Number of sample areas	Housing units eligible[1]	
		Interviewed	Not interviewed
1996 to 2001	754	46,800	3,200
1995	792	56,700	3,300
1990 to 1994	729	57,400	2,600
1989	729	53,600	2,500
1986 to 1988	729	57,000	2,500
1985	[2]629/729	57,000	2,500
1982 to 1984	629	59,000	2,500
1980 to 1981	629	65,500	3,000
1977 to 1979	614	55,000	3,000
1973 to 1976	461	46,500	2,500
1972	449	45,000	2,000
1968 to 1971	449	48,000	2,000

[1]Excludes about 2,500 Hispanic households added from the previous November sample. (See "CPS March Supplement.")

[2]The Census Bureau redesigned the CPS following the 1980 Decennial Census of Population and Housing. During phase-in of the new design, housing units from the new and old designs were in the sample.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

The estimates in this report for 1983 (from March 1984 CPS) and later also employ a revised survey weighting procedure for people of Hispanic origin. In previous years, we inflated weighted sample results to independent estimates of the noninstitutional population by age, sex, and race. There was no specific control of the survey estimates for the Hispanic population. Since then, the Census Bureau developed independent population controls for the Hispanic population by sex and detailed age groups. Revised weighting procedures incorporate these new controls. The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of error: nonsampling and sampling. The accuracy of an estimate depends on both types of error, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in

[1]Hispanics may be of any race. This report shows information on the Hispanic population collected in the 50 states and the District of Columbia, and therefore, does not include residents of Puerto Rico.

responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling variability. We can attribute nonsampling errors to several sources including the following:

- Inability to obtain information about all cases in the sample.
- Definitional difficulties.
- Differences in the interpretation of questions.
- Respondents' inability or unwillingness to provide correct information.
- Respondents' inability to recall information.
- Errors made in data collection, such as in recording or coding the data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

CPS undercoverage results from missed housing units and missed people within sample households. Compared with the level of the 1990 decennial census, overall CPS undercoverage is about 8 percent. Undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. As described previously, ratio estimation to independent age-sex-race-Hispanic population controls partially corrects for the bias due to undercoverage. However, biases exist in the estimates to the extent that missed people in missed households or missed people in interviewed households have different characteristics from those of interviewed people in the same age-sex-race-Hispanic group.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table E–2 shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month, but these are a typical set of coverage ratios.

Answers to questions about money income often depend on the memory or knowledge of one person in a household. Recall problems can cause underestimates of income in survey data, because it is easy to forget minor or irregular sources of income. Respondents may also misunderstand what the Census Bureau considers money income or may simply be unwilling to answer these questions correctly because the questions are considered too personal. See Appendix C, Current Population Reports, Series P60-184, *Money Income of Households, Families, and Persons in the United States: 1992* for more details.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978 and Technical Paper 63, *The Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce.

Comparability of data. Data obtained from the CPS and other sources are not entirely comparable. This results from differences in field representative training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Use caution when comparing results from different sources.

Table E–2. **March CPS Coverage Ratios**

Age	Non-Black		Black		All races		
	Male	Female	Male	Female	Male	Female	Total
0 to 14 years	0.942	0.951	0.880	0.904	0.932	0.943	0.937
15 to 19 years	0.864	0.910	0.885	0.751	0.867	0.884	0.876
20 to 24 years	0.823	0.877	0.707	0.757	0.808	0.859	0.834
25 to 29 years	0.863	0.919	0.755	0.810	0.850	0.903	0.877
30 to 34 years	0.880	0.950	0.671	0.833	0.855	0.934	0.895
35 to 44 years	0.899	0.940	0.684	0.863	0.875	0.930	0.903
45 to 54 years	0.938	0.961	0.778	0.953	0.923	0.960	0.942
55 to 64 years	0.932	0.953	0.834	0.929	0.923	0.951	0.938
65 to 74 years	0.932	0.977	0.939	0.958	0.932	0.975	0.956
75 years and older	1.019	1.008	0.910	0.961	1.011	1.004	1.007
15 years and older	0.902	0.945	0.767	0.858	0.887	0.934	0.912
0 years and older	0.911	0.946	0.802	0.871	0.898	0.936	0.917

U.S. Census Bureau, Demographic Statistical Methods Division.

Table E–3. **CPS Standard Error Parameters for Income and Nonincome Characteristics: 2000**

Characteristics	Total or White		Black		Hispanic	
	a	b	a	b	a	b
ALL INCOME LEVELS						
People						
Total	-0.000011	2,454	-0.000108	2,810	-0.000120	2,810
Male	-0.000024	2,454	-0.000239	2,810	-0.000237	2,810
Female	–0.000022	2,454	-0.000196	2,810	-0.000231	2,810
Age						
15 to 24	-0.000063	2,454	-0.000481	2,810	-0.000328	2,810
25 to 44	-0.000030	2,454	-0.000261	2,810	-0.000184	2,810
45 to 64	-0.000040	2,454	-0.000421	2,810	-0.000188	2,810
65 and over	-0.000075	2,454	-0.001007	2,810	-0.000536	2,810
Households, Families, and Unrelated Individuals						
Total	-0.000010	2,241	-0.000094	2,447	-0.000104	2,447
Households with children under age 18	-0.000010	2,241	-0.000094	2,447	-0.000104	2,447
NONINCOME CHARACTERISTICS						
People						
Employment status	-0.000008	1,586	-0.000154	3,296	-0.000187	3,296
Educational attainment	-0.000011	2,369	-0.000103	2,680	-0.000077	1,811
Marital Status, Household and Family Characteristics						
Some household members	-0.000019	5,211	-0.000209	7,486	-0.000221	7,486
All household members	-0.000023	6,332	-0.000309	11,039	-0.000327	11,039
Households, Families, and Unrelated Individuals						
Total	-0.000010	2,068	-0.000072	1,871	-0.000080	1,871

Notes: To obtain parameters prior to 2000, multiply by the appropriate factor in Table E–4.

For nonmetropolitan residence categories multiply the a and b parameters by 1.5.

For foreign-born and noncitizen characteristics for Total and White, multiply the a and b parameters by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

For regional estimates, multiply the a and b parameters by 0.85, 1.03, 1.08, and 1.09 for Northeast, Midwest, South, and West, respectively.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The Bureau of Labor Statistics redesigned the questionnaire to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The Census Bureau modified the March supplemental income questions for adaptation to computer-assisted interviewing, but did not change definitions and concepts. Because of these and other changes, one should use caution when comparing estimates from data collected in 1994 or later years with estimates from earlier years.

Data users should also use caution when comparing CPS estimates in this report, which reflect 1990 census-based population controls, with estimates for 1992 (from March 1993 CPS) and earlier years, which reflect 1980 census-based population controls. This change in population controls had relatively little impact on summary measures, such as means, medians, and percentage distributions, but did have a significant impact on levels. For example, 1990-based population controls caused a 1-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1994 and later years will differ from those published for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

Since the Census Bureau did not use independent population control totals for people of Hispanic origin before 1983, compare Hispanic estimates over time cautiously.

Based on the results of each decennial census, the Census Bureau gradually introduces a new sample design for the CPS. During this phase-in period, the Census Bureau collects CPS data from sample designs based on different censuses. While most CPS estimates have been unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing estimates across years for metropolitan/nonmetropolitan categories. For more information, see Appendix C, Current Population Reports, Series P60-193, *Money Income in the United States: 1995 (With Separate Data on Valuation of Noncash Benefits).*

Note when using small estimates. The Census Bureau shows summary measures (such as medians, means, and percentage distributions) only when the base is 75,000 or greater. Because of the large standard errors involved, summary measures would probably not reveal useful information when computed on a smaller base. However, we display estimated numbers even though the relative standard errors of these numbers are larger than those for corresponding percentages. These smaller estimates permit combinations of the categories to suit data users' needs. Take care in the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Estimation of median incomes. The Census Bureau has changed the methodology for computing median income over the past few years. The Census Bureau has computed medians using either Pareto interpolation or linear interpolation. Currently, we are using linear interpolation to estimate all medians. Pareto interpolation assumes a decreasing density of population within an income interval; whereas, linear interpolation assumes a constant density of population within an income interval. The Census Bureau calculates estimates of median income for 1979 through 1987 and associated standard errors using Pareto interpolation if the estimate is larger than $20,000 for people or $40,000 for families and households. This is because the width of the income interval containing the estimate is greater than $2,500.

We calculated estimates of median income for 1976, 1977, and 1978 and associated standard errors using Pareto interpolation if the estimate was larger than $12,000 for people or $18,000 for families and households. This is because the width of the income interval containing the estimate is greater than $1,000. We calculated all other estimates of median income and associated standard errors for 1976 through 2000 and almost all of the estimates of median income and associated standard errors for 1975 and earlier using linear interpolation.

Thus, use caution when comparing median incomes above $12,000 for people or $18,000 for families and households for different years. Median incomes below those levels are more comparable from year to year since they have always been calculated using linear interpolation. For an indication of the comparability of medians calculated using Pareto interpolation with medians calculated using linear interpolation, see Series P-60, No. 114, *Money Income in 1976 of Families and Persons in the United States.*

Sampling variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated by methods described in "Standard errors and their use," are primarily measures of sampling variability, but they may include some nonsampling error.

Standard errors and their use. Data users must use a number of approximations to derive, at a moderate cost, standard errors applicable to all the estimates in this report. Instead of providing an individual standard error for each estimate, we have provided two parameters, a and b, to calculate standard errors for each type of characteristic.

Table E–3 has CPS standard error parameters for various types of characteristics. Table E–4 provides factors to approximate CPS standard error parameters for estimates before 2000. Table E–5 provides CPS Hispanic parameters for estimates before 1984. Table E–6 provides CPS Asian and Pacific Islander parameters for income and nonincome characteristics. Table E–7 has the year-to-year CPS correlation coefficients for income characteristics.

The sample estimate and its standard error enable one to construct a confidence interval, a range that would include the average result of all possible samples with a known probability. For example, if all possible

Table E–4. **CPS Factors to Apply to a and b Parameters for Estimates Prior to 2000**

Characteristic	Factor
NON-HISPANIC	
1995 to 1999	1.00
1989 to 1994	0.92
1988	1.02
1981 to 1987	0.86
1967 to 1980	0.75
HISPANIC	
1995 to 1999	1.00
1989 to 1994	0.92
1988	1.19
1984 to 1987	0.86

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table E–5. **CPS Standard Error Parameters for Income and Nonincome Characteristics of Hispanics: 1972 to 1983**

Characteristics	1972 - 1980		1981 - 1983	
	a	b	a	b
ALL INCOME LEVELS				
People				
Total	-0.000020	3,000	-0.000301	3,357
Male	-0.000043	3,000	-0.000615	3,357
Female	-0.000038	3,000	-0.000591	3,357
Age				
15 to 24	-0.000080	3,000	-0.000961	3,357
25 to 44	-0.000065	3,000	-0.000668	3,357
45 to 64	-0.000077	3,000	-0.001459	3,357
65 and over	-0.000147	3,000	-0.004124	3,357
Farm	(X)	(X)	(X)	(X)
Households, Families, and Unrelated Individuals				
Total	-0.000014	2,420	-0.000237	2,708
Farm	(X)	(X)	(X)	(X)
Households with children under age 18	-0.000014	2,420	-0.000237	2,708
NONINCOME CHARACTERISTICS				
People				
Employment status	(X)	(X)	(X)	(X)
Educational attainment	-0.000015	2,344	-0.000152	2,623
Farm	(X)	(X)	(X)	(X)
Total, Marital Status, Other				
Some household members	-0.000026	5,069	-0.000294	5,673
All household members	-0.000044	10,199	-0.000592	11,414
Households, Families, and Unrelated Individuals				
Total	-0.000020	1,626	-0.000022	1,820
Farm	(X)	(X)	(X)	(X)

X Not applicable.

Note: Data users should multiply the a and b parameters by 1.5 for nonmetropolitan residence categories. The Census Bureau did not publish income data for Hispanics before 1972.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Data users may also use standard errors to perform hypothesis testing; this is a procedure for distinguishing between population parameters using sample estimates. One common type of hypothesis appearing in this report is that two population parameters are different. An example of this would be comparing the median annual income of Black families with the median annual income of White non-Hispanic families.

One can perform tests at various levels of significance. The significance level of a test is the probability of concluding that the characteristics are different when, in fact, they are the same. All statements of comparison in the text were tested at the 0.10 level of significance or better. This means that the absolute value of the estimated difference between characteristics is greater than or equal to 1.645 times the standard error of the difference.

Table E–6. **CPS Standard Error Parameters for Income and Nonincome Characteristics of Asians and Pacific Islanders and American Indians and Alaskan Natives: 2000**

Characteristics	a	b
ALL INCOME LEVELS		
People	-0.000267	2,810
Households, families, and unrelated individuals	-0.000232	2,447
NONINCOME CHARACTERISTICS		
People		
Marital status, household and family characteristics:		
Some household members	-0.000533	7,486
All household members	-0.000786	11,039
Households, families, and unrelated individuals	**-0.000177**	**1,871**

Note: To obtain parameters prior to 2000, multiply by the appropriate factor in Table E–4. Income data for Asians and Pacific Islanders and American Indians and Alaskan Natives were not collected prior to 1988.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table E–7. **CPS Year-to-Year Correlation Coefficients for Income Estimates: 1960 to 2000**

Characteristics	People	Families, households, and unrelated individuals
Total	**0.30**	**0.35**
White	0.30	0.35
Black	0.30	0.35
Other races	0.30	0.35
Hispanic[1]	0.45	0.55

[1]Hispanics may be of any race.

Note: These correlations are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlations are zero. For Asians and Pacific Islanders and American Indians and Alaskan Natives, use the correlation coefficient for total.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Standard errors of estimated numbers. The approximate standard error, s_x, of an estimated number shown in this report can be obtained using the formula

$$s_x = \sqrt{ax^2 + bx} \qquad (1)$$

Here x is the size of the estimate and a and b are the parameters in Table E–3 or E–6 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the set of parameters for the characteristic which will give the largest standard error.

Illustration. There were 72,375,000 family households in 2001. Use the appropriate parameters from Table E–3 and formula (1) to get

Estimate, x	72,375,000
a parameter	-0.000010
b parameter	2,068
Standard error	312,000
90% confidence interval	71,862,000 to 72,888,000

The standard error is calculated as

$$s_x = \sqrt{(-0.000010)\,(72{,}375{,}000)^2 + (2{,}068)\,(72{,}375{,}000)} = 312{,}000$$

The 90-percent confidence interval for the estimated number of family households in 2001 is calculated as 72,375,000 ± 1.645 x 312,000.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard errors of estimated percentages. The reliability of an estimated percentage, computed using sample data from both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table E–3 or E–6 indicated by the numerator.

The approximate standard error, $s_{x,p}$, of an estimated percentage is approximately equal to

$$s_{x,p} = \sqrt{\frac{b}{x}\,p(100 - p)} \qquad (2)$$

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \le p \le 100$), and b is the parameter in Table E–3 or E–6 associated with the characteristic in the numerator of the percentage.

Illustration. There were 12,525,000 or 17.3 percent of the 72,375,000 family households maintained by female householders with no husband present. Use the appropriate parameter from Table E–3 and formula (2) to get

Estimate, p	17.3
Base, x	72,375,000
b parameter	2,068
Standard error	0.2
90% confidence interval	17.0 to 17.6

The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{2{,}068}{72{,}375{,}000}\ (17.3)\ (100.0 - 17.3)} = 0.2$$

The 90-percent confidence interval for the estimated percentage of family households that were maintained by female householders with no husband present is calculated as 17.3 ± 1.645 x 0.2.

Standard error of a difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x-y} = \sqrt{s_x^{\,2} + s_y^{\,2} - 2rs_xs_y} \qquad (3)$$

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. Table E–7 contains the correlation coefficient, r, for year-to-year comparisons for CPS income estimates of numbers and proportions. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration. The median earnings of all male full-time, year-round workers in 2000, x, was $37,339 and the median earnings of all female full-time, year-round workers in 2000, y, was $27,355. The apparent difference between the median income of males and females in 2000 was $9,984. The approximate standard errors, s_x and s_y, are $137 and $107, respectively. Use formula (3) with r = 0 to get

	x	y	difference
Estimate	$37,339	$27,355	$9,984
Standard error	$137	$107	$174
90% confidence interval	$37,114 to $37,564	$27,179 to $27,531	$9,698 to $10,270

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{(137)^2 + (107)^2} = 174$$

The 90-percent confidence interval for the estimated difference between the median income of male and female full-time, year-round workers in 2000 is calculated as $9,984 ± 1.645 x $174. Because this interval does not contain zero, we can conclude with 90-percent confidence that the median income of male full-time, year-round workers in 2000 was larger than the median income of female full-time, year-round workers in 2000.

Standard error of a ratio. Certain estimates may be calculated as the ratio of two numbers. Compute the standard error of a ratio, x/y, using

$$s_{x/y} = \frac{x}{y}\sqrt{\left[\frac{s_x}{x}\right]^2 + \left[\frac{s_y}{y}\right]^2 - 2r\frac{s_xs_y}{xy}} \qquad (4)$$

Calculate the standard error of the numerator, s_x, and that of the denominator, s_y, using formulas described earlier.

In formula (4), r represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of people in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r. An example of this type is the mean number of children per family with children.

For all other types of ratios, r is assumed to be zero. If r is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio. Examples of this type are the mean number of children per family and the family poverty rate.

NOTE: For estimates expressed as the ratio of x per 100 y or x per 1,000 y, multiply formula (4) by 100 or 1,000, respectively, to obtain the standard error.

Illustration. The median earnings for full-time, year-round female workers in 2000, x, was $27,355 and the median earnings for full-time, year-round male workers in 2000, y, was $37,339. The ratio of the median earnings is 0.73. The approximate standard errors, s_x and

s_y, are \$107 and \$137, respectively. Using formula (4) with r = 0 to get

	x	y	ratio
Estimate	\$27,355	\$37,339	.73
Standard error	\$107	\$137	.0039
90% confidence interval	\$27,179 to \$27,531	\$37,114 to \$37,564	0.72 to 0.74

The standard error is calculated as

$$s_{x/y} = \frac{27{,}355}{37{,}339}\sqrt{\left[\frac{107}{27{,}355}\right]^2 + \left[\frac{137}{37{,}339}\right]^2} = .0039$$

The 90-percent confidence interval for the ratio of the median earnings for full-time, year-round female workers to the median earnings for full-time, year-round male workers is calculated as $.73 \pm 1.645 \times 0.0039$.

Standard errors of other estimates. This report provides standard errors for most estimates in the respective tables, or includes a formula showing how to calculate them. For information on calculating other standard errors, contact Aneesah Stephenson at e-mail address: dsmd_s&a@ccmail.census.gov.